

A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus

Roman Biek^{1*}, J. Caroline Henderson^{1,5}, Lance A. Waller¹, Charles E. Rupprecht⁵, and Leslie A. Real¹

¹Department of Biology and Center for Disease Ecology, Emory University, 1510 Clifton Road, Atlanta, GA 30322; ⁵Rabies Section, Centers for Disease Control and Prevention, 1600 Clifton Road, Mail-stop G33, Atlanta, GA 30333; and ¹Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Road, Atlanta, GA 30322

Edited by John C. Avise, University of California, Irvine, CA, and approved March 26, 2007 (received for review January 26, 2007)

Emerging pathogens potentially undergo rapid evolution while expanding in population size and geographic range during the course of invasion, yet it is generally difficult to demonstrate how these processes interact. Our analysis of a 30-yr data set covering a large-scale rabies virus outbreak among North American raccoons reveals the long lasting effect of the initial infection wave in determining how viral populations are genetically structured in space. We further find that coalescent-based estimates derived from the genetic data yielded an amazingly accurate reconstruction of the known spatial and demographic dynamics of the virus over time. Our study demonstrates the combined evolutionary and population dynamic processes characterizing the spread of pathogen after its introduction into a fully susceptible host population. Furthermore, the results provide important insights regarding the spatial scale of rabies persistence and validate the use of coalescent approaches for uncovering even relatively complex population histories. Such approaches will be of increasing relevance for understanding the epidemiology of emerging zoonotic diseases in a landscape context.

invasion | phylogeography | *Procyon lotor* | wildlife disease | zoonoses

Infectious diseases emerging in wildlife populations, and the concomitant threats to humans and domestic animals, continue to be a source of great public concern (1). Such emergences are often characterized by the spatial incursion of a pathogen into a completely susceptible host population (2–4). Opportunities to reduce or control the rate of spatial spread in wild animals are, unfortunately, usually quite limited. However, the lack of intervention measures also means that these pathogen invasions can be viewed as large-scale natural experiments that can reveal much about the spatial, ecological, and genetic processes simultaneously unfolding during population expansions, yielding insights not only relevant to disease emergence but also more generally to invasion biology and the reconstruction of biogeographical histories from molecular data.

RNA viruses offer particularly good opportunities for the study of pathogen invasion because they exhibit ecological and evolutionary dynamics that occur on similar time scales (4–6). Rabies is the most important viral zoonosis worldwide, likely causing >50,000 human deaths annually (7). Besides bats, terrestrial carnivore species, including raccoons (*Procyon lotor*), are the most important rabies reservoirs. Although raccoons are common throughout North America, their significance as a primary rabies host before the 1970's was limited to the southeastern U.S., particularly Florida (8). This localized focus changed dramatically during 1977, when a raccoon-specific rabies virus variant (RRV) was detected in West Virginia, several hundred km north of the previously identified focus of raccoon rabies. Radiating from this location, RRV spread explosively along the mid-Atlantic coast in the following years, and, by 1999, the epizootic affected an area of many thousand square kilometers (9).

In this article, we present a spatiogenetic analysis of this 30-yr rabies virus epizootic among North American raccoons, which

for most of its duration went on uncontrolled and which from its index case in 1977 to the present has been documented in unusually fine temporal and spatial detail (9–14). In addition to the documented case data, genetic viral samples from this outbreak have been stockpiled by the Centers for Disease Control and Prevention since 1982. Together, these data offer a rare chance to examine how the demographic and spatial processes of spread and population expansion over 30 yr have shaped viral evolution on a landscape scale. Furthermore, the exact knowledge about how the epizootic unfolded in time and space provides an opportunity to test how accurately the documented population history of raccoon rabies can be reconstructed from genetic data by using available analytical tools.

Results

Based on phylogenetic analysis, we were able to distinguish seven genetic lineages of RRV based on an operational definition (see *Materials and Methods*). All but two lineages represented groups with posterior values >80%; the two remaining lineages received low to moderate statistical support (57 and 69%, respectively) but were supported in their distinction also on the basis of spatial clustering of viruses (Fig. 1A). Each of the seven lineages exhibited a general direction of spread relative to the first reported cases in Pembleton county, WV: Spread toward the southeast, southwest, and east was each associated with one particular group, whereas spread in the northwestern and northeastern direction was each marked by two different groups (Fig. 1A and B).

Counties sampled 5–14 yr or even 15–25 yr after they had experienced their first raccoon rabies cases consistently yielded members of the same genetic lineages that had colonized that area initially (Fig. 1C). A correlation of spatial and temporal distances derived from the phylogeny revealed the different rates of spread at different times during the epizootic: Compared with the initial wave when infections were estimated to have spread on average 38.4 ± 3.8 km/yr, viruses sampled at later stages indicated a mean spread rate of 9.5 ± 1.4 km/yr ($P < 0.001$).

Of 324 inferred substitutions, 273 (84%) occurred in putatively neutrally evolving sites (third codon position or noncoding sequence). All remaining substitutions were unique to one or few

Author contributions: R.B. and L.A.R. designed research; R.B. and J.C.H. performed research; L.A.W. and C.E.R. contributed new reagents/analytic tools; R.B. and L.A.W. analyzed data; and R.B., C.E.R., and L.A.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: HPD, highest posterior density; R_e , effective reproductive number; RRV, raccoon rabies virus.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ886039–DQ886076, DQ888332–DQ888369, and EF508133–EF508144).

*To whom correspondence should be addressed. E-mail: rbiek@emory.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/070741104/DC1.

© 2007 by The National Academy of Sciences of the USA

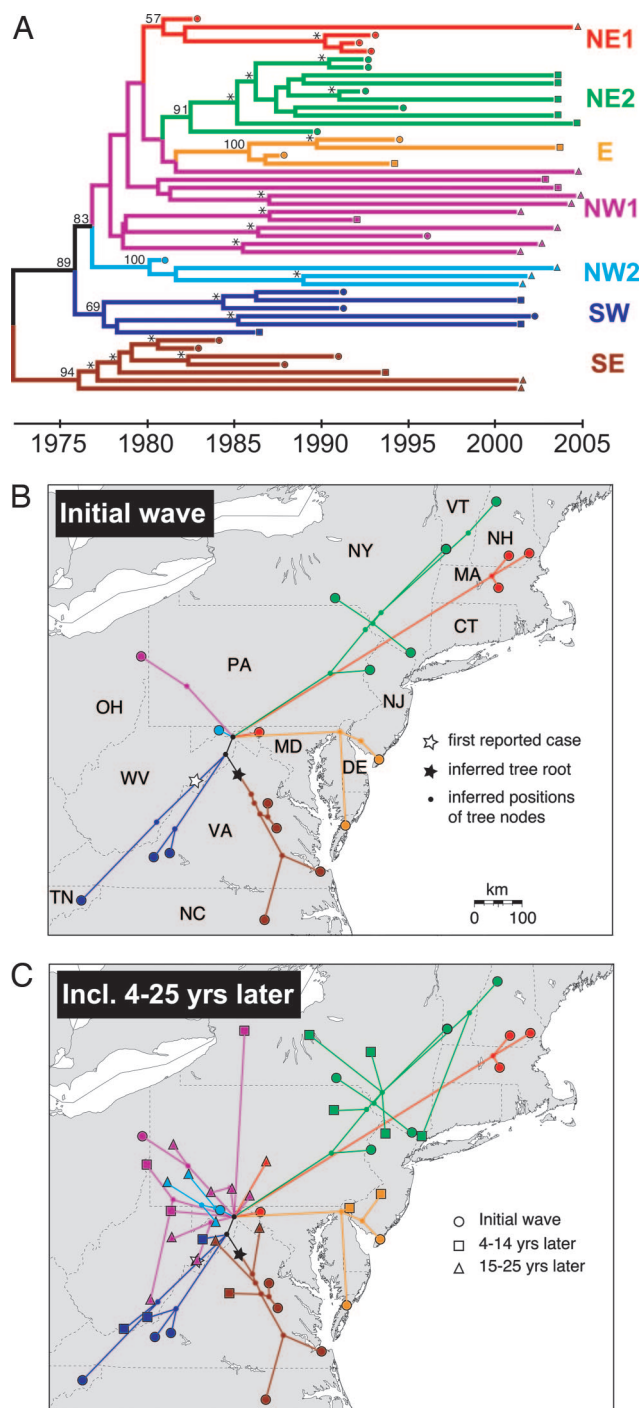


Fig. 1. Phylogenetic relationships among virus samples collected during a RRV epizootic in eastern North America. (A) Maximum clade credibility (MCC) tree from Bayesian coalescent analysis (32). Color-coding of branches distinguishes seven lineages (see *Materials and Methods*). Clade credibility values $>50\%$ are shown up to the lineage level only; nodes within lineages that received $>70\%$ posterior support are indicated by asterisks. (B) Portion of MCC tree that corresponds to initial infection wave projected onto the landscape. Tree tips represent centroids of counties that were sampled during the initial wave (i.e., collected within 37 months of the first reported case of RRV in that county). Latitude and longitude of nodes (including root, shown as black star) were estimated by using phylogenetic generalized least squares (37). White star marks Pendleton County, WV, where the epizootic's first case was reported in 1977. (C) Full MCC tree projected onto landscape, including samples collected 4–14 (squares) and 15–25 yr (triangles) after the first county case. Maps were produced with the help of online map creator (available at www.aquarius.geomar.de/omc/).

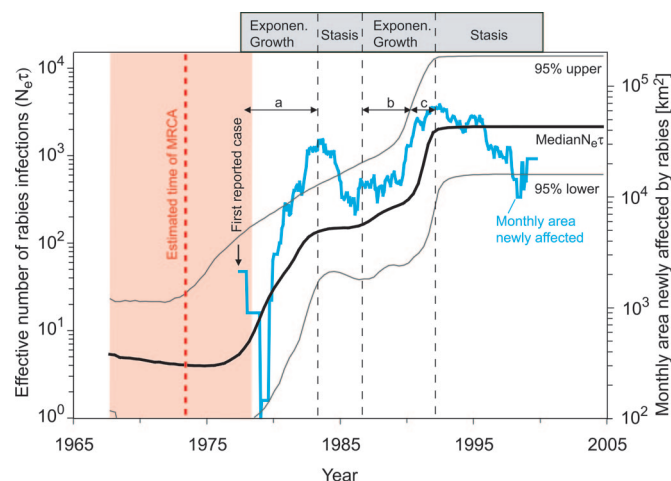


Fig. 2. Number of RRV infections, 1977–2005, estimated from genetic and case data. Median effective number of RRV infections (thick black line) was estimated by using a Bayesian skyline plot and represents the product of effective population size (N_e) and generation time (τ) in years. Thin black lines represent 95% highest posterior density (HPD) intervals. Estimated time associated with the most recent common ancestor is indicated by red dashed line; HPD interval is shown as pink shaded area. Light blue line represents the 15-months moving average of the monthly area (in km²) newly affected by RRV between 1977 and 1999 as an index of the number of rabid raccoons through time. Exponen., exponential.

sequences, corroborating earlier findings that genetic diversification of rabies virus is probably not driven by the selection of novel variants (15).

The Bayesian estimation in BEAST yielded an evolutionary rate of 2.9×10^{-4} substitutions/site/yr for the concatenated sequence data. Results of the relaxed clock method indicated that systematic changes in rate variation over the course of the epizootic were negligible [see [supporting information \(SI\) Fig. 4](#)], thus demonstrating the validity of the strict clock approach. In conjunction with the evolutionary rate, we identified the year 1973 (highest posterior density interval, HPD: 1967–1978) as the estimated date of the most recent common ancestor for the mid-Atlantic epizootic (Fig. 2). This date was consistent with the date of the first documented cases in 1977, which fell within the HPD limits of the coalescent estimate and therefore reaffirmed that the use of a strict molecular clock resulted in reasonable temporal estimates that fit the known epizootic history.

Our coalescent-based reconstruction of the demographic history of the mid-Atlantic epizootic (Fig. 2) suggested that the effective number of infected raccoon had not increased at a uniform rate but instead had occurred in three periods of exponential growth, interspersed with periods of stable population size: After an initial rapid exponential increase (1977–1983), the effective number of infections reached a short plateau (1983–1986), followed by first moderate (1986–1990), then rapid exponential growth (1990–1992), before reaching another plateau (1992–1999).

Our epidemiological index for the number of infected raccoons through time, the cumulative area of all counties newly affected by rabies each month, showed very similar temporal dynamics to those identified in the coalescent analysis (Fig. 2). After the first reported case in 1977, the newly affected area rose more or less continuously until ≈ 1983 , when the trend was reversed and the virus invaded new area at a declining rate. Around 1986, this negative trend stopped and, by 1990, changed again to a steady increase that lasted until 1992. Between 1992 and 1998 the size of the new area fairly consistently declined again but showed one final spiked increase during 1999. The fact

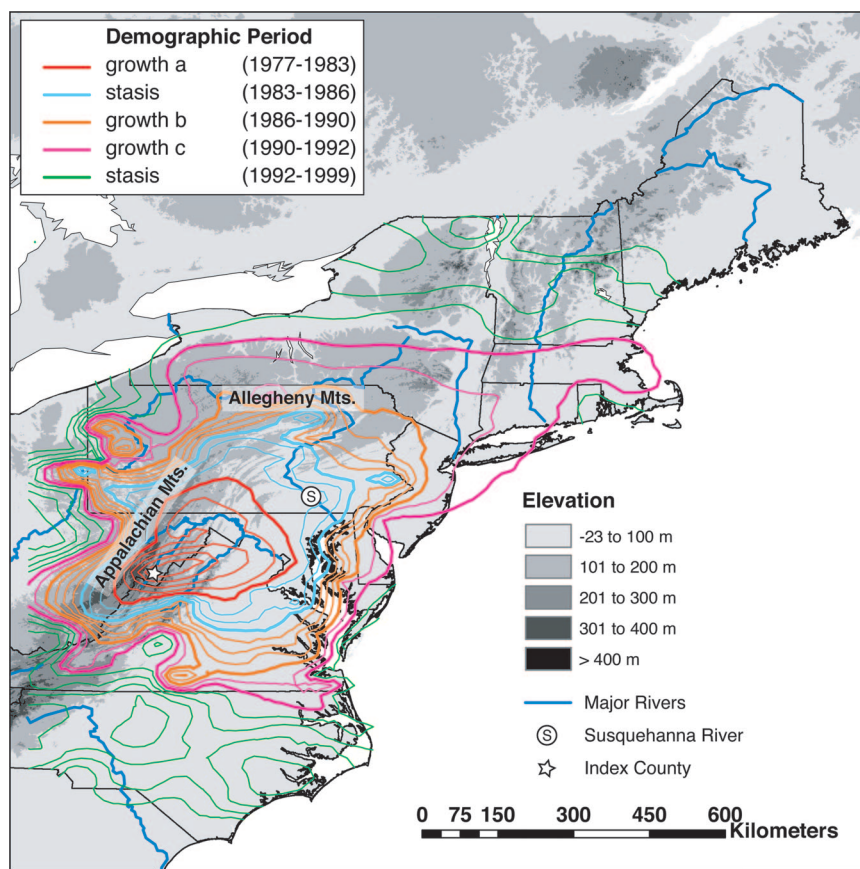


Fig. 3. Annual rate of RRV spread along the U.S. mid-Atlantic relative to elevation and major rivers, 1977–1999. Annual contours of spread were computed based on the date of the first case of RRV reported in a county by using a kriging routine available in ArcGIS Geostatistical Analyst (42). Colors of contours distinguish periods of fast and slow viral population growth (see Fig. 2) that were identified from coalescent demographic analysis. Thicker lines indicate the respective endpoints of demographic periods.

that such a tight correspondence between the epidemiological and the coalescent data were seen for the timing of demographic periods starting and ending further validates the accuracy of the molecular clock rate estimate.

Annual contours for the RRV invasion revealed high levels of spatial heterogeneity in the expansion process (Fig. 3). For example, RRV spread into the areas north and east of the index case at a variable but generally faster rate compared with areas in the south. By 1992, RRV in the north had reached New Hampshire, >700 km from the index case, whereas expansion toward the south had only progressed as far as the North Carolina coast (≈ 300 km). The reasons for this difference are unclear but may be related to different environmental conditions and lower raccoon densities from Virginia southwards (16). Westward expansion of RRV was particularly slow, likely because of the Appalachian mountain ranges forming a barrier to raccoon movement. The same was true, albeit to a lesser degree, for northward spread across the Allegheny mountains. Conversely, areas of low elevation along the Atlantic coast, especially in the states of New Jersey, New York, Connecticut, and Massachusetts, were associated with the fastest annual rates of RRV spread (Fig. 3).

Splitting the genetic data according to direction of expansion and repeating the coalescent analysis demonstrated that the distinction of several demographic periods was not possible for viruses spreading south. Instead, the estimated number of infected raccoons for these lineages (SE and SW) increased at a slow but exponential rate at first but then changed to stable population sizes by the late 1980s (SI Fig. 5). In contrast, the

same distinct demographic periods identified from the full data set were also all visible when limiting the analysis to the northern lineages only (E, NE1, NE2, NW1, NW2; SI Fig. 5). Thus, consistent with the observation of more homogeneous spread toward the south, the demographic fluctuations we identified in the Bayesian skyline analysis (Fig. 2) solely pertained to viral spread in northern directions.

Results of the Bayesian skyline plot (Fig. 2) indicated that the mid-Atlantic rabies epizootic contained three distinct periods of exponential population growth (see above). Hence, we used the three corresponding point estimates of population growth rate to determine R_e , the effective reproductive rate, of raccoon rabies. Depending on the specific growth period and the assumed length of generation time, estimates of R_e ranged from 1.02 to 1.16 (Table 1).

Discussion

The combination of rapid evolution and spatial diffusion of RRV after its introduction to the mid-Atlantic region resulted in a distinct phylogeographic pattern, in which each viral lineage became associated with a specific radial direction of rabies spread from the epizootic origin. Interestingly, this spatial segregation of lineages was not only evident from viruses sampled during the initial infection wave (Fig. 1B) but even remained true for samples collected from counties that had experienced their first raccoon rabies case 5–25 yr earlier (Fig. 1C). Previous analysis of RRV dynamics had shown that counties reexperience rabies outbreaks every 4 yr on average (9). If these subsequent outbreaks, which tend to have smaller amplitudes

Table 1. Bayesian point estimates of population growth rates (r) and corresponding estimates of the effective reproductive number (R_e) for raccoon rabies

Exponential growth period [†]	Years	r	R_e	
			$\tau = 1$ mo	$\tau = 3$ mo
a	1977–1983	0.53	1.04	1.13
b	1986–1990	0.26	1.02	1.07
c	1990–1992	0.63	1.05	1.16

τ , assumed average generation time (in months) between rabies virus infections.

[†]Demographic periods as shown in Fig. 2.

(9), generated their own traveling infection waves, viral lineages should, over time, become admixed over increasingly larger areas. That this was not the case thus indicates that later waves remain much more local in scale, as also supported by our calculation that infections at this stage spread over average distances that were $\approx 75\%$ smaller than during the initial spread. These results imply that RRV in its enzootic stage is maintained at relatively small spatial scales and that local persistence does not depend upon regular immigration of infected individuals.

Our observations fit theoretical expectations for expanding populations, in particular those of a “surfing mutation” model (17, 18). According to this model, mutations (in our case represented by viral lineages) that arise near the front of an expansion wave have the chance to be carried by the wave over long distances and to thereby reach high overall frequencies. In contrast, mutations or lineages arising at a later stage, and thus behind the expansion wave, lack the same opportunity to be dispersed. Our study presents an empirical example for this phenomenon, which so far has only been shown in simulation studies and which has important implications for reconstructing historic population expansions from molecular data (17, 18).

According to our analysis, all viruses sampled from the mid-Atlantic epizootic can be traced back to a common ancestor in the year 1973. Although the 4-yr time lag between our point estimate and the observed emergence in 1977 may be explained by sampling variation, it could also reflect epidemiological processes. One possible interpretation is that the mid-Atlantic outbreak resulted from the simultaneous introduction of several closely related viruses in 1977. Alternatively, and in our view more likely, RRV may have already been present in West Virginia for several years before its detection. Such time lags are characteristic for species invasions (19) and seem particularly probable for rabies, given that the reproductive number of the virus in raccoons appears to be small (see below) and that detection probability will be low during times when the number of infected individuals is still limited. We noted a similar time lag between local introduction and emergence in another recent RRV outbreak (J.C.H., R.B., C.E.R., and L.A.R., unpublished work) and thus believe that it may be a potentially common feature of rabies epizootics.

The demographic history of RRV inferred from sequence data showed a strikingly close correspondence to the actually observed population dynamics of the virus (Fig. 2). Importantly, this not only confirmed the existence of several distinct demographic periods during the RRV expansion but also their timing and different growth rates. RRV population growth was only seen at times during which the size of the newly invaded area continued to increase. This underscores the degree to which overall dynamics of the virus depend on processes at the wave front and suggests that behind this front, raccoon mortality due to RRV quickly limits the number of susceptibles available for infection. Despite the widespread use of coalescent-based methods in molecular epidemiology in recent years, it is rarely

possible to validate the obtained estimates through independent data. Our results therefore provide a powerful illustration of the detailed level of information these approaches can extract from adequate molecular data.

What could have caused the number of infected raccoons to change so dramatically at different times during the epizootic? Landscape features can have pronounced effects on the rate of local rabies spread (10, 11) and may therefore also affect viral dynamics on larger scales. Because our genetic inference methods integrated over data from the entire RRV population, we could only expect to see the effect of large physical features that altered the dynamics of several viral lineages simultaneously. We noted a significantly lower rate of RRV spread from the index case southwards compared with the rate of viral invasion toward the north (Fig. 3), and therefore conducted separate demographic analyses for northern and southern lineages. Results indicated that the series of distinct demographic periods (Fig. 2) only applied to lineages spreading northward, whereas southern lineages exhibited much simpler dynamics (*SI Text*). Consequently, only physical features encountered during northward spread had the potential to induce changes in the number of infected raccoons.

Continuous mountain ranges, which constitute poor raccoon habitat at higher elevations and probably less favored routes during dispersal, had a pronounced effect on RRV spread (Fig. 3). The Appalachian mountain chains in particular, which were not crossed during the first growth period (1977–1983), clearly played an important role in limiting the westward spread of RRV (16). Similarly, the Allegheny Mountains appear to have slowed down the spatial expansion of RRV to the north. Because the Susquehanna River and the Chesapeake Bay simultaneously limited viral spread to the east, we would argue that the first static demographic period (1983–1986) may be attributable to the virus becoming confined by these physical barriers. Notably, the period of fastest growth, 1990–1992, coincided with rapid viral invasion into low-elevation areas along the Atlantic coast, which are also highly urbanized and support some of the highest raccoon densities (20, 21). The incursion of RRV into these areas therefore would be expected to strongly affect the number of infections. The final demographic plateau in RRV infections reached in 1992–1999 can be explained by the combined effects of additional physical barriers (i.e., Great Lakes, Atlantic Ocean) and poorer raccoon habitat at higher latitudes. In addition, the distribution of an oral rabies vaccine along the western and northern wave front, which began during the mid-1990s (22), likely contributed to stabilizing the number of infections at that point. Interestingly, we saw no obvious effect of large rivers, with the possible exception of the aforementioned Susquehanna River, on wave velocity and the inferred demographic periods. Large rivers have been shown to significantly reduce the rate of local RRV spread (10, 11), but our results suggest that these effects may become less important on larger scales.

The estimates we obtained for the effective reproductive ratio (R_e) of RRV were low, ranging only from 1.02 to 1.16 (Table 1). Compared with R_0 , which describes the expected average number of secondary cases resulting from a single infection, R_e represents the realized equivalent, which will always be lower than R_0 because it reflects, among other factors, the variance in the number of secondary cases and the local depletion of susceptible hosts (23). Indeed, estimates of R_0 for rabies in domestic dogs and wildlife have generally been >2.0 (24, 25). Despite the expectation of a lower rate, we were surprised that R_e just barely exceeded 1.0, the value required for an epizootic to persist. A low reproductive ratio makes pathogen dynamics more susceptible to stochastic fluctuations during periods of low population size. As discussed above, this may explain the lag time between the estimated age of the most recent common ancestor of the mid-Atlantic focus and the first detection of rabid

raccoons (Fig. 2). In addition, it could also have significant implications for attempts to model emergence dynamics of RRV and other rabies variants by using deterministic approaches because predictions derived from these models may fail to accurately capture population processes that are driven by stochastic effects.

Our work underscores the tight linkage among the spatial, demographic, and genetic processes taking place during species invasion over a novel landscape and supports the notion that evolutionary events occurring during an initial expansion have profound and lasting effects on the genetic population structure of an invader (17, 18). In many cases, this initial geographic structure may become later obscured by subsequent movement and population turnover, making it difficult to correctly infer increasingly complex population histories from genetic data (26). Even after ≈ 30 yr, however, this has not been the case for RRV, for which phylogeographic structure remains largely intact. If the same holds true for other viral zoonoses, it may be possible in many cases to recover the origin and initial direction of viral spread through retrospective analysis, even if genetic sampling is undertaken decades later. Especially in situations where surveillance data are unavailable or inadequate, this approach could provide critical insights regarding landscape epidemiology and the historic factors involved in local disease emergence.

This study demonstrates that coalescent methods are capable of obtaining detailed and accurate accounts of nonlinear population dynamics in the past, and illustrates how these dynamics may be affected by the local physical environment. Similar approaches, and a further integration of genetic and spatial data, are likely to become increasingly important for epidemiological studies of infectious diseases on natural landscapes.

Materials and Methods

Rabies Virus Sampling. We examined brain tissues of 44 rabid raccoons collected between 1982 and 2004 that had been submitted to the Rabies Section of the U.S. Centers for Disease Control and Prevention (CDC) for diagnostic confirmation and variant typing by monoclonal antibodies. These samples were selected from a much larger set of samples collected as part of regular rabies surveillance carried out by state and local public health departments and the U.S. Department of Agriculture Wildlife Services according to the following criteria: First, we included all available samples that had been collected within 37 months after that county had experienced its first reported case of raccoon rabies ($n = 17$). Because most counties did not produce secondary outbreaks within this time period (9), these samples are considered to be representative of the initial wave of rabies infection. We also included published sequences from raccoons ($n = 3$) that had been sampled within this time frame (27). In addition, we examined 27 samples that had been collected 52–302 months after a county had experienced its first reported case and thus had the potential to yield virus introduced from other areas by subsequent infection waves. Samples had associated information regarding the date and location (county) of sampling; the county centroid was considered a sample's spatial origin (see SI Table 2).

Sequence Amplification. We used RT-PCR to amplify the complete rabies nucleoprotein (*N*) gene, part of a noncoding region immediately following the 3' end of *N*, and a large portion of the glycoprotein (*G*) gene (see SI Text). Final data sets were trimmed to 1,359 bp for *G*, 1365 bp for *N*, and 87 bp for noncoding sequence immediately following *N*. A partition homogeneity test (28) conducted in Paup* (29) revealed no significant incongruence between the *G* and *N* data ($P = 0.89$); therefore, sequences were concatenated for analysis. Sequences have been submitted to GenBank under accession nos.

DQ886039–DQ886076, DQ888332–DQ888369, and EF508133–EF508144.

Genetic Coalescent Analysis. We used Akaike's Information Criterion (AIC) to select an appropriate model of molecular evolution from a broad range of candidate models (see SI Text) (30, 31). The best model included different base frequencies and substitution parameters for codon positions one and two vs. position three (HKYuf₁₁₂ + CP₁₁₂ + G₁₁₂). We estimated evolutionary rates and demographic history from raccoon rabies sequences in the software application Beast 1.4, which implements a Bayesian coalescent approach by using Markov-Chain Monte Carlo sampling (32). Among other parameters, the program yields estimates of the evolutionary rate if supplied with genetic sequences sampled at different points in time and if temporal separation between samples is long relative to the frequency of mutational events (33, 34).

We estimated the viral effective population size through time (interpretable as the product of the effective number of infected raccoons and the virus generation time τ) by using a Bayesian skyline plot (35), which is also part of the Beast program. Instead of fitting the sequence data to a demographic model specified *a priori* (i.e., logistic growth, exponential growth, etc.), the Bayesian skyline plot estimates the population dynamics in the past from the posterior distribution of effective population size estimates, by assuming a highly parametric stepwise model of population size change, and can thus account for more complex demographic scenarios. The number of potential population size transitions must be chosen beforehand, and, in our case, we chose to allow 15 steps in the underlying demographic model. We also repeated the same analysis after splitting the data set into a southern group (containing all sequences of the SE and SW lineages) and a northern group (all other sequences) motivated by earlier observations that dynamics of spread from the index case had differed for these general directions (16).

Analyses in Beast were based on 22 million steps, of which the first 2 million were discarded as a burn-in period. This length resulted in effective sampling sizes of at least 350 for all estimated parameters. Samples of trees and parameters were recorded every 1,000 steps. To assess whether the assumption of a strict molecular clock was justified and whether our results were sensitive to variation in the clock rate, we repeated our coalescent analysis in Beast by using a relaxed clock model (36).

Based on the maximum clade credibility (MCC) tree found in the Beast analysis, we estimated the latitude and longitude associated with each supported tree node (defined as posterior support $>70\%$), including the tree root, by using the phylogenetic generalized least-squares method (37) available in program Compare 4.6 (38) as described (6). For each supported branch, we divided the geographic distance between the two nodes by the estimated branch length in years (again from MCC tree), to obtain a distribution of distances over which infections would have been able to spread. We distinguished between branches that corresponded to the initial expansion wave (see *Rabies Virus Sampling* above) and branches that reflected mutations that arose after the expansion wave had already moved through an area. Three internal branches, which could not be readily assigned to either group, were excluded from the analysis.

To distinguish specific lineages in the phylogeny, we used an operational definition according to which a lineage (*i*) had to correspond to a supported (but potentially paraphyletic) group, (*ii*) had to contain at least four sequences, and (*iii*) was not subdivided further if this would have resulted in "left-over" groups containing fewer than four sequences.

Ancestral reconstruction, as implemented in program BASEML of the PAML package (39), was used to infer the number of substitutions at each site in the viral genetic sequence along the MCC tree.

Demographic Analysis of Case Data. Temporal data specifying the first month of occurrence of raccoon rabies in a county were available from July 1977 to October 1999. These data were used to determine whether genetic samples could be regarded as representing the initial infection wave or some later time point (see *Rabies Virus Sampling* above).

A local raccoon population that recently experienced a rabies outbreak will take, on average, 4 yr to regain population sizes sufficient to support a secondary outbreak, which will usually be of smaller amplitude (9). Thus, the majority of infected raccoons at any given time should be found in completely susceptible populations that experience rabies infection for the very first time. These populations will be located at the perimeter of the range of the epizootic, i.e., at the wave front. Based on this rationale, and making the simplifying assumptions of homogeneous raccoon densities and infection rates, we derived a monthly index for the number of infected raccoons by calculating the cumulative area of all counties experiencing rabies for the first time that month. To dampen the effects of heterogeneity in the data caused by temporal and sampling effects, we used a moving average with a window size of 15 months.

Estimating the Effective Reproductive Number (R_e) of Raccoon Rabies.

The basic reproductive number (R_0) is defined as the average number of secondary cases resulting from a single infection in a completely susceptible host population. Persistence of a pathogen within a host population therefore requires an $R_0 \geq 1$. R_0 can

be calculated in a straightforward manner (40) by using $R_0 = r\tau + 1$, where r is the growth rate in the number of infections and τ represents the average time between infections. Because we derive our estimate for r from genetic data, it is most appropriately thought of as a measure of increase in the effective number of infections, analogous to the concept of effective population size in population genetics. We therefore refer to our estimate for the reproductive number as R_e , rather than R_0 . We obtained point estimates of r by visually determining start and end points for each of the three demographic phases within the population size curve and calculating the slope of a linear fit for each of the three periods. For the parameter τ , we assumed values of 1–3 mo (see *SI Text*).

Effect of Landscape on Rabies Spread. Spatial kriging using ArcGIS Geostatistical Analyst (41) was used to compute annual contours of rabies spread throughout the mid-Atlantic region based on the date of the first reported case for each county. This contour map was then overlaid in ArcGIS with an elevation model and the distribution of major rivers.

We thank S. Nadin-Davis and R. J. Rudd for assistance, two anonymous reviewers for excellent suggestions, and colleagues in the health departments throughout the U.S. for generation of rabies surveillance data. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of their institutions. This research was supported through National Institutes of Health Grant R01-AI047498 and U.S. Department of Agriculture Grant 0371004129CA.

- Daszak P, Cunningham AA, Hyatt AD (2000) *Science* 287:443–449.
- Swinton J, Harwood J, Grenfell BT, Gilligan CA (1998) *J Anim Ecol* 67:54–68.
- Fischer JR, Stallknecht DE, Luttrell P, Dhondt AA, Converse KA (1997) *Emerg Infect Dis* 3:69–72.
- Real LA, Henderson JC, Biek R, Snaman J, Jack TL, Childs JE, Stahl E, Waller LA, Tinline R, Nadin-Davis S (2005) *Proc Natl Acad Sci USA* 102:12107–12111.
- Holmes EC (2004) *Mol Ecol* 13:745–756.
- Biek R, Drummond AJ, Poss M (2006) *Science* 311:538–541.
- World Health Organization (2005) *WHO Tech Rep Ser* 931:1–88.
- Rupprecht CE, Smith JS (1994) *Semin Virol* 5:155–164.
- Childs JE, Curns AT, Dey ME, Real LA, Feinstein L, Bjornstad ON, Krebs JW (2000) *Proc Natl Acad Sci USA* 97:13666–13671.
- Smith DL, Lucey B, Waller LA, Childs JE, Real LA (2002) *Proc Natl Acad Sci USA* 99:3668–3672.
- Lucey BT, Russell CA, Smith DL, Wilson ML, Long A, Waller LA, Childs JE, Real LA (2002) *Vector Borne Zoonotic Dis* 2:77–86.
- Real LA, Russell CA, Waller LA, Smith DL, Childs JE (2005) *J Hered* 96:253–260.
- Russell CA, Smith DL, Waller LA, Childs JE, Real LA (2004) *Proc Biol Sci* 271:21–25.
- Smith DL, Waller LA, Russell CA, Childs JE, Real LA (2005) *Prev Vet Med* 71:225–240.
- Holmes EC, Woelk CH, Kassis R, Bourhy H (2002) *Virology* 292:247–257.
- Childs JE, Curns AT, Dey ME, Real LA, Rupprecht CE, Krebs JW (2001) *Vector Borne Zoonotic Dis* 1:253–267.
- Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) *Proc Natl Acad Sci USA* 101:975–979.
- Klopfstein S, Currat M, Excoffier L (2006) *Mol Biol Evol* 23:482–490.
- Shigesada N, Kawasaki K (1997) *Biological Invasions: Theory and Practice* (Oxford Univ Press, Oxford).
- Jones ME, Curns AT, Krebs JW, Childs JE (2003) *J Wildl Dis* 39:869–874.
- Riley SPD, Hadidian J, Manski DA (1998) *Can J Zool* 76:1153–1164.
- Slate D, Rupprecht CE, Rooney JA, Donovan D, Lein DH, Chipman RB (2005) *Virus Res* 111:68–76.
- Keeling MJ, Grenfell BT (2000) *J Theor Biol* 203:51–61.
- Coleman PG, Dye C (1996) *Vaccine* 14:185–186.
- Haydon DT, Randall DA, Matthews L, Knobel DL, Tallents LA, Gravenor MB, Williams SD, Pollinger JP, Cleaveland S, Woolhouse ME, et al. (2006) *Nature* 443:692–695.
- Avise JC (2000) *Phylogeography: The History and Formation of Species* (Harvard Univ Press, Cambridge, MA).
- Nadin-Davis SA, Huang W, Wandeler AI (1996) *J Virol Methods* 57:1–14.
- Farris JS, Kallersjo M, Kluge AG, Bult C (1994) *Cladistics* 10:315–319.
- Swofford DL (2003) *Paup*4.0b10* (Sinauer Associates, Sunderland, MA).
- Posada D, Crandall KA (1998) *Bioinformatics* 14:817–818.
- Shapiro B, Rambaut A, Drummond AJ (2006) *Mol Biol Evol* 23:7–9.
- Drummond AJ, Rambaut A (2003) *BEAST v1.4*, <http://evolve.zoo.ox.ac.uk/beast/>.
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) *Trends Ecol Evol* 18:481–488.
- Rambaut A (2000) *Bioinformatics* 16:395–399.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) *Mol Biol Evol* 22:1185–1192.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) *PLoS Biol* 4:e88.
- Martins EP, Hansen TF (1997) *Am Nat* 149:646–667.
- Martins EP (2004) *COMPARE, version 4.6*. (Department of Biology, Indiana University, Bloomington IN) <http://compare.bio.indiana.edu/>.
- Yang Z (1997) *Comput Appl Biosci* 13:555–556.
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH (2001) *Science* 292:2323–2325.
- Johnston K, Ver Hoef JM, Krivoruchko K, Lucas N (2001) *Using ArcGIS Geostatistical Analysis* (ESRI, Redlands, CA).